International Journal of Computer Science and Information Security (IJCSIS),
Vol. 16, No. 4, April 2018

Kinect Sensor based Indian Sign Language Detection with Voice Extraction

Shubham Juneja[1] , Chhaya Chandra[2] , P.D Mahapatra[3], Siddhi Sathe[4], Nilesh B. Bahadure[5] and Sankalp Verma[6]

[1]Material Science Program, M.Tech first year student, Indian Institute of Technology
Kanpur, UttarPradesh, India
junejashubh@gmail.com
[2]B.E, Electrical and Electronics Engineering, Bhilai Institute of Technology
Raipur, Chhattisgarh, India
chhaya.chandra02@gmail.com
[3]B.E, Electrical and Electronics Engineering, Bhilai Institute of Technology
Raipur, Chhattisgarh, India
pushpadas27495@gmail.com
[4]Department of Electrical & Electronics Engineering, B.E. Final Year Student, Bhilai Institute of Technology
Raipur, Chhattisgarh, India
sathesiddhi1996@gmail.com
[5]Assoc. Professor at Department of Electronics & Telecommunication Engineering, MIT College of Railway Engineering & Research
Solapur, Maharashtra, India
nbahadure@gmail.com
[6]Assoc. Professor at Department of Electrical & Electronics Engineering, Bhilai Institute of Technology
Raipur, Chhattisgarh, India
sankalpverma99@gmail.com

Abstract

We are progressing towards new discoveries and inventions in the field of science and technology, but unfortunately, very rare inventions could have helped the problems faced by the physically challenged people who face difficulties in communicating with normal people as they use sign language as their prime medium for communication. Mostly, the sign languages are not understood by the common people. Studies say that many research works have been done to eliminate such kind of communication barrier. But those work involves the functioning of Microcontrollers or by some other complicated techniques. Our study advances this process by using the Kinect sensor. Kinect sensor is a highly sensitive motion sensing device with many other applications. Our workflow from capturing of an image of the body to conversion into the skeletal image and from image processing to feature extraction of the detected image hence getting an output along with its meaning and voice. The experimental results of our proposed algorithm are also very promising with an accuracy of 94.5%.

Keywords: *Hidden Markov Model (HMM), Image Processing, Kinect Sensor, Skeletal Image*

1. Introduction

Since a very long time, we are experiencing a better life due to the existence of various electronic systems and the sensing elements almost in every field.Physically challenged people find it easier to communicate with each other and common people using different sets of hand gestures and body movements. We hereby provide an aid to very efficiently express themselves in front of common people wherein their sign languages will be automatically converted into text and speeches. Their hand and body gestures will be taken as inputs by the sensor, making it easier for them to understand.

This is a machine to human interaction system which includes Kinect sensor and Matlab for processing the data given as input.

There have been multitudinous researches done till date, but this paper provides a direct and flexible system for deaf and dumb people. It extracts voice from the human gesture of sign language as well as generates images and texts depending upon the input gestures given to the system. The very first step is to give the input as gesture data to the Kinect sensor, by this it senses the data and a 3-D image are created. This data is then transferred to Matlab where it is interfaced through the programming along with image processing and feature extraction using different segmentations and Hidden Markov Model (HMM) algorithm. From the complete segmented body, only the image of the hand is cropped, the gesture of that hand is then equated with the available image in the database and if they match the speech and text is obtained as output making it easier for the

common people to understand it. By this, the disabled people will be confident enough to express their views anywhere and everywhere despite physically challenged.

2. Literature Review

The Sign Language detection is considered an efficient way by which physically challenged people can communicate. Many researchers have studied and investigations are done on different algorithms to make the process easier.

Gunasekaran and Manikandan [1] have worked on a technique using PIC Microcontroller for detection of sign languages. The authors stated that their method is better as it solves the real time problems faced by the disabled ones. Their work involves extraction of voice as soon as any sign language is detected.

Kiratey Patil et al. [2] worked on detection of American Sign Language. Their work is based on accessing American Sign Language and converting into English and the output flashes on LCD. This way their work may omit the communication gap between common and disabled ones.

Tavari et al. [3] worked on recognition of Indian Sign languages by hand gestures. They proposed an idea of recognizing images formed by different hand movement gestures. They proposed an idea of recognizing images formed by different hand movement gestures. They used a web camera in their work. For identifying signs and translation of text to voice, Artificial Neural Network has been used.

Simon Lang [4] worked on Sign Language detection. He proposed a system that uses Kinect sensor instead of a web camera. Out of nine signs performed by many people, 97% detection rate have been seen for eight signs. The important body parts are sensed by Kinect sensor easily and Markov Model is used continuously for detection.

Sign Writing system proposed by Cayley et al. [5] deals with procuring in helping deaf people by using stylus and screen contraption for the written literacy in Sign Language. They have provided databases for enhancing the studies in another paper so that the sequence of the characters can be stored and retrieved in order to signify the sign language and then the editing could be done. In order to enhance their work on sensing algorithm, they are further researching on it.

According to Singha and Das [6], several Indian sign languages have been acknowledged by the process of skin filtering, hand cropping feature extraction and classification by making use of Eigen value weighted Euclidean distance. Hence out of 26 alphabets, only dynamic alphabets 'H & J' were not taken into account & they will be considered in their future studies.

According to Xiujuan Chaivfgtxtgf.,njoh et al. [7] for hard and body tracking 3-D motion by using Kinect Sensor is more effective and clear. This makes sign language detection easier.

According to our earlier work [8], the efficiency was 92% but now our efficiency has increased to 94.5%. We have used very simple algorithm here rather than using FCM.

From the above studies, it has been observed that few methods are only proposed for hand gestures recognition and few are only for feature extraction. Also from the above done survey, it is understood that no precise idea about feature extraction in easiest way is mentioned. But this problem is solved in our study. We have proposed an algorithm and used HMM technique also. Gestures are identified easily, that information is then matched with our preset databases and voice is extracted. This process enables the common people to understand the sign language easily.

3. Methodology

We have proposed an algorithm which follows the following steps:

1. After the detection of the body in front of the Microsoft X-box Kinect Sensor as shown in Fig.1, it locates the joints of the body by pointing it out and hence we get a skeletal image.

Fig. 1 Kinect Sensor

2. Then the segmented image of the body is formed from the skeletal image. The area of the hands where the signs are captured are cropped out of the whole segmented image. Then the cropped image is converted into dots and dashes. The length of the dash is 4 unit and the spacing between the two lines of dashes is also 4 units.
3. Through observation, we found that wherever the length of the dashes is greater than 4 units it resembles the image of the cropped hand.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

4. In our proposed algorithm, we have taken the concept of loops, it is used to detect the black points that are the space between the dashes. This detection of black points determines the position of dashes by successive subtraction of points in the iterations which is going on and on.
5. The basic algorithm behind this work is that after successive subtraction of points if the value is equal to 4 then there is no data and if the value is greater than 4 then there is the actual image of the cropped hand.
6. Now, here arises a problem that how we detect the location of fingers. So the algorithm behind this is based on the formation of matrices of the black points detected earlier. In an iteration, the matrix coordinates of the finger are four times the number of lines of dashes.
7. To highlight the fingers, we plotted star point of the same coordinates as of the fingers and for more precise feature extraction the image processing is followed by filtration of the image that means the star points plotted on the figures go through following conditions.
    - If they fall on a straight line either horizontal or vertical.
    - If they fall on a constant slope.
    - If they fall within 4-unit coordinate difference.
8. Then it eliminates the identical points which we call as garbage point. Now we get the filtered image but the process of feature extraction continues for identification of fingers that whether it is an index finger, middle finger, ring finger, little finger or thumb.

   Following is the Table 1 which shows the range of coordinates in which the fingers are detected.

Table 1: Range of coordinates of fingers

S No.	Finger	Range of coordinates
1	Index	80-88
2	Middle	60-68
3	Ring	44-52
4	Little	32-40
5	Thumb	104-112

The flow chart of the above algorithm is shown in following Fig.2:

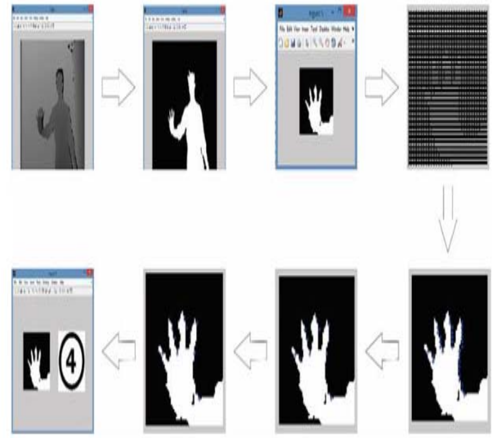

Fig. 2 Flow chart of the proposed algorithm

4. Hidden Markov Model

HMM [5], [9] is the algorithm which says that the actual stages of the work continued in the system is not visible the final output after the whole processing is only visible.

HMM works on probability and it uses a hidden variable of any input data and select them for various observations and then process all those variables through Markov process. HMM undergoes four stage process:

- Filtering: This state involves the computation which takes place during the hidden process of the given statistical parameters.
- Smoothing: This state does the same work as the filtering process but works in between the sequence wherever needed.
- Most likely Explanation: This state is different from the above two states. It is generally used whenever HMM is exposed to a different number of problems and to find overall maximum possible state sequences.
- Statistical Significance: This state of HMM is used to obtain statistical data and evaluate the data of the possible outcome.

5. Result

Finally, after the detection of the whole image of fingers or we can say that a complete hand ANDing operation continues in Matlab for the final output.

 https://sites.google.com/site/ijcsis/
ISSN 1947-5500

The detected image of the sign is searched in the database for its meaning and as and when the match is found search is complete and we get the final output along with the image of the meaning of sign and its voice.

We can take the example as, if number 4 is to be detected by the Kinect sensor then the person gestures 4 using his hands. The Kinect captures the skeletal image of the body as shown in the Fig. 3.

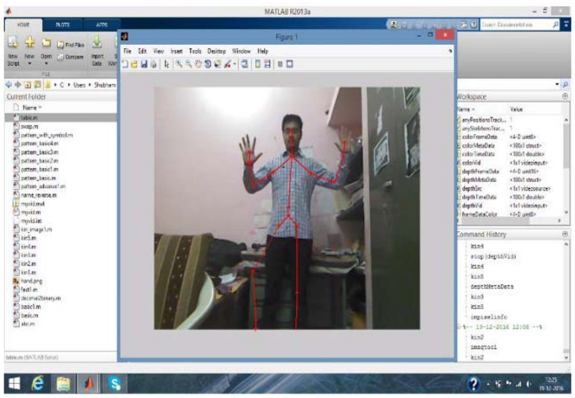

Fig. 3 Skeletal image

After skeletal image, the image is converted into depth image as in Fig. 4.

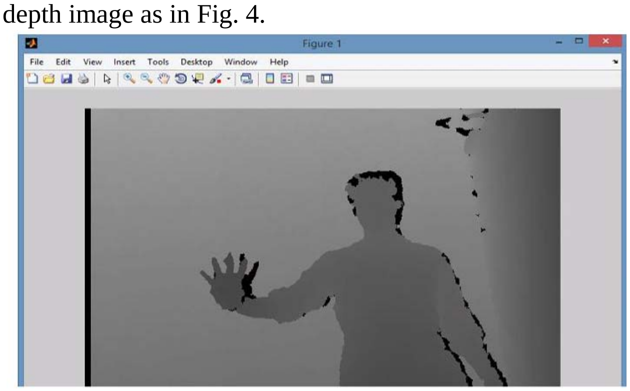

Fig. 4 Depth image

Then the image is converted into its segmented image as shown in Fig. 5.

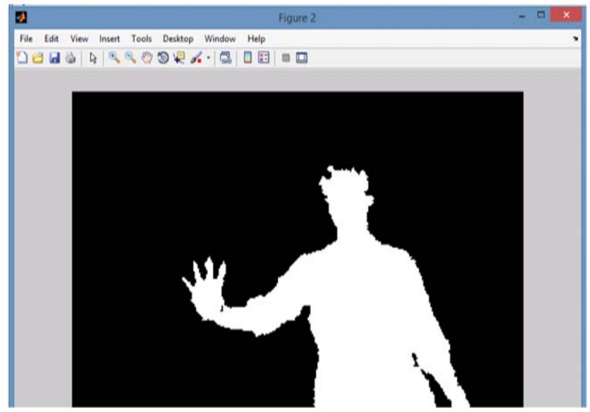

Fig. 5 Segmented image

The image of the hand is cropped from the segmented image as shown in Fig. 6.

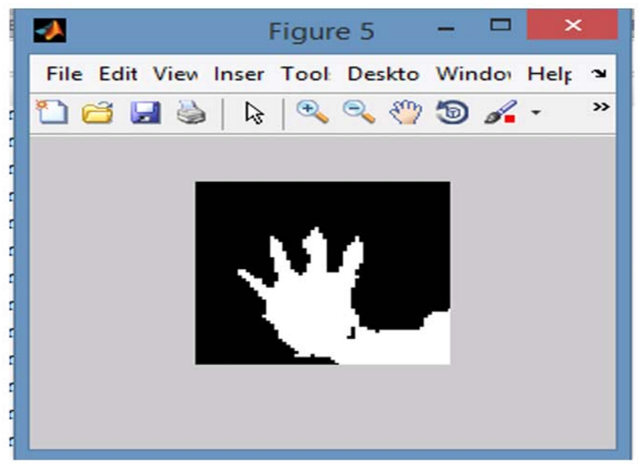

Fig. 6 Cropped image

The cropped image is then converted into a figure with dots and dashes as shown in Fig. 7.

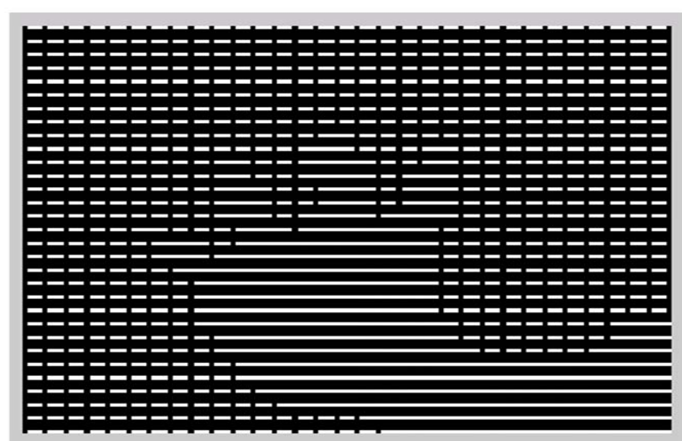

Fig. 7 Image with dots and dashes

The filtration of the figure after star marking it to detect the fingers is done in 3 steps as shown in Fig. 8.

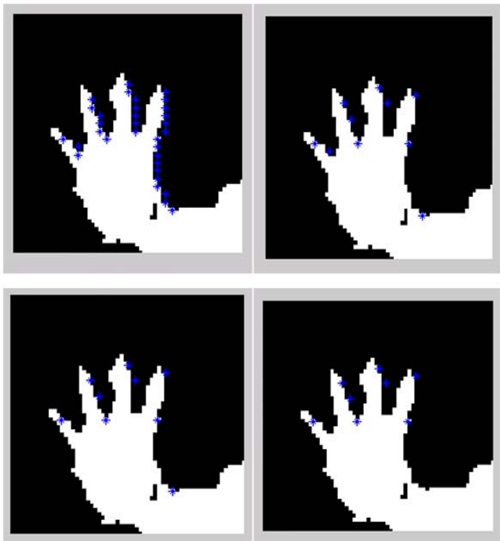

Fig. 8 Filtration of figure after star marking

The detailed information of the fingers detected are shown in command window which is shown in Fig. 9.

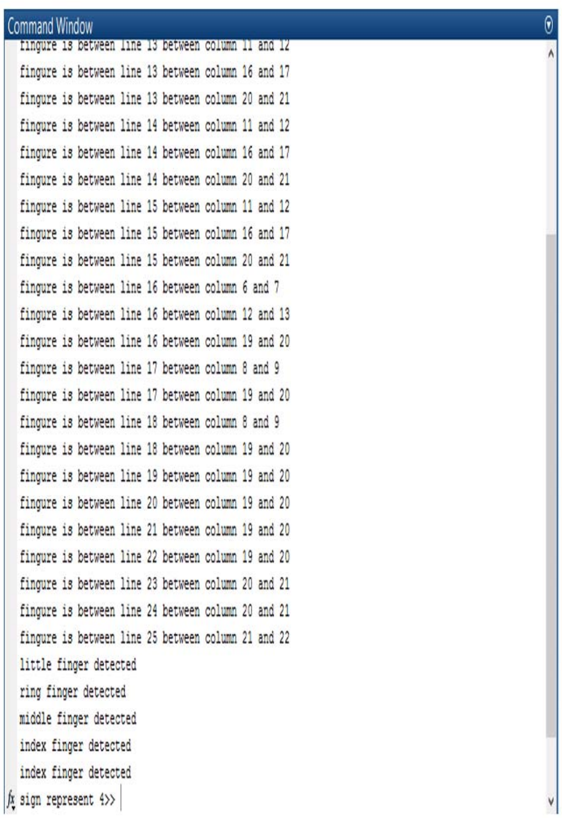

Fig. 9 Detailed information of fingers detected

After the filtration is done the database is searched for its match and hence we get an output in the form of image as shown in Fig. 10(a) and voice which is plotted in the form of the histogram as shown in Fig. 10(b).

Fig. 10(a) Output in the form of image

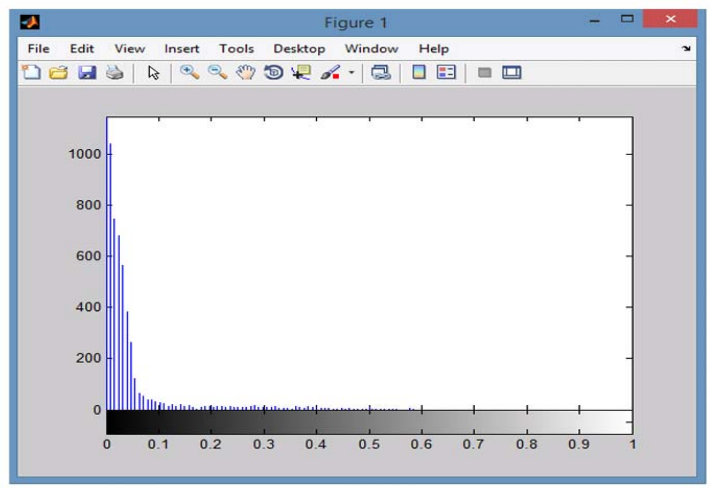

Fig. 10(b) Voice in the form of histogram

Overall output of various input given to the system shown in Fig.11.

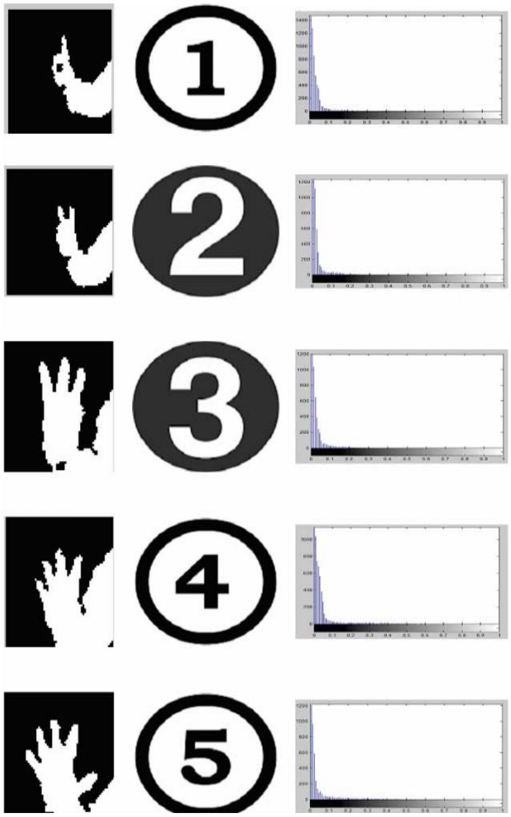

Fig. 11 [a] Cropped images, [b] Image of the meaning of signs, [c] Histogram image of voice output

Detailed analysis of the various outputs with respect to given inputs have been tabulated in Table 2.

Table 2: Detailed Analysis

S No.	No. of correct attempts	No. of wrong attempts	Accuracy(%)
1	50	0	100
2	50	0	100
3	49	1	96
4	48	2	92
5	49	1	96
6	49	1	96
7	47	3	88
8	47	3	88

Total no. of attempts = 50

Accuracy = (No. of correct attempts – No. of wrong attempts)/Total no. of attempts

Hence we get the total accuracy as 94.5%.

6. Conclusions and Future work

With references to all the earlier studies, our work provides the better accessibility with the simpler algorithm and more precise output.

Since programming is done for the detection of left hand the coordinates are taken accordingly. Our algorithm gives all the relevant information about the coordinates of each and every finger detected.

This system is very flexible and user-friendly as the user can be of any age, gender, size or color, the results will be same. But the intensity of light and distance of the body from the sensor affects the efficiency. For working effectively with the devices it is suggested to keep the Kinect sensor at a height of about 62 cm from the ground and the body to be detected should be distanced at about 90cm.

In our earlier work [8] the exact location of the fingers and the detailed information about them has not been determined, so we have overcome these problems here.

The star points that we have marked earlier is not completely filtered, very few points still remain there. Sometimes misinterpretation of detected fingers also occurs but its possibility is one out of ten.

Acknowledgments

We would like to extend our gratitude and sincere thanks to all people who have taken a great deal of interest in our work and helped us with a valuable suggestion.

References

[1] K. Gunasekaran and R. Manikandan, “Sign language to speech translation system using pic microcontroller”, International Journal of Engineering and Technology, vol. 05, no. 02, pp. 1024–1028, April 2013.

[2] K. Patil, G. Pendharkar, and G. N. Gaikwad, “American sign language detection”, International Journal of Scientific and Research Publications, vol. 04, pp. 01–06, November 2014.

[3] N. V. Tavari, A. V. Deorankar, and P. N. Chatur, “Hand gesture recognition of Indian sign language to aid physically impaired people”, International Journal of Engineering Research and Applications, pp. 60–66, April 2014.

[4] S. Lang, “Sign language recognition with Kinect”, Master’s thesis, Freie University Berlin, 2011.

[5] C.-Y. Kao and C.-S. Fahn, “A human-machine interaction technique: hand gesture recognition based on Hidden Markov Models with trajectory of hand motion”, Elsevier Advanced in Control Engineering and Information Science, vol. 15, pp. 3739–3743, 2011.

[6] J. Singha and K. Das, “Indian sign language recognition using Eigen- value weighted Euclidian distance based classification technique”, International Journal of Advanced Computer Science and Applications, vol. 04, no. 02, pp. 188–195, 2013.

[7] X. Chai, G. Li, Y. Lin, Z. Xu, Y. Tang, X. Chen, and M. Zhou, “Sign language recognition and translation with Kinect”, in IEEE International Conference on Automatic Face and Gesture Recognition, April 2013.

[8] N.B.Bahadure, Sankalp Verma, Shubham Juneja, Chhaya Chandra, D.K.Mishra, and P.D. Mahapatra, “Sign Language Detection with Voice Extraction in Matlab using Kinect Sensor”, in International Journal of Computer Science and Information Security, vol.14, no. 12, pp.858-863, December 2016.

[9] T. Starner and A. Pentland, “Visual recognition of American sign language using Hidden Markov models”, in Proceedings of International workshop face and gesture recognition, 1995, pp. 189–194.

[10] S. R. Ganapathy, B. Aravind, B. Keerthana, and M. Sivagami, “Conversation of sign language to speech with human gestures”, in Elsevier 2nd International Symposium on Big Data and Cloud Computing, vol. 50, 2015, pp. 10–15.

[11] M. Boulares and M. Jemni, “3d motion trajectory analysis approach to improve sign language 3d-based content recognition”, in Elsevier Proceedings of International Neural Network Society Winter Conference, vol. 13, 2012, pp. 133–143.

[12] D. Vinson, R. L. Thompson, R. Skinner, and G. Vigliocco, “A faster path between meaning and form? Iconicity facilitates sign recognition and production in British sign language”, Elsevier Journal of Memory and Language, vol. 82, pp. 56–85, March 2015.

[13] K. Cormier, A. Schembri, D. Winson, and E. Orfanidou, “A faster path between meaning and form? Iconicity facilitates sign recognition and production in British sign language”, Elsevier Journal of Cognition, vol. 124, pp. 50–65, May 2012.

[14] H. Anupreethi and S. Vijayakumar, “Msp430 based sign language recognizer for dumb patient”, in Elsevier International Conference on Modeling Optimization and Computing, vol. 38, 2012, pp. 1374–1380.

[15] C. Guimaraes, J. F. Guardez, and S. Fernanades, “Sign language writing acquisition- technology for writing system”, in IEEE Hawaii International Conference on System Science, 2014, pp. 120–129.
[16] C. Guimaraes, J. F. Guardezi, S. Fernanades, and L. E. Oliveira, “Deaf culture and sign language writing system- a database for a new approach to writing system recognition technology”, in IEEE Hawaii International Conference on System Science, 2014, pp. 3368–3377.

[17] J. S. R. Jang, C. T. Sun, and E. Mizutani, Neuro - Fuzzy and Soft Computing. Eastern Economy Edition Prentice Hall of India, 2014.

[18] S. C. Pandey and P. K. Misra, “Modified memory convergence with fuzzy PSO”, in Proceedings of the World Congress on Engineering, vol. 01, 2 - 4 July 2007.

[19] M. S. Chafi, M.-R. Akbarzadeh-T, M. Moavenian, and M. Ziejewski, “Agent based soft computing approach for component fault detection and isolation of CNC x - axis drive system”, in ASME International Mechanical Engineering Congress and Exposition, Seattle, Washington, USA, November 2007, pp. 1–10.

[20] S. C. Pandey and P. K. Misra, “Memory convergence and optimization with fuzzy PSO and ACS”, Journal of Computer Science, vol. 04, no. 02, pp. 139–147, February 2008.

[21] N. B. Bahadure, A. K. Ray, and H. P. Thethi, “Performance analysis of image segmentation using watershed algorithm, fuzzy c – means of clustering algorithm and Simulink design”, in IEEE International Conference on Computing for Sustainable Global Development, New Delhi, India, March 2016, pp. 30–34.

[22] H.-D. Yang, “Sign language recognition with Kinect sensor based on conditional random fields”, Sensors Journal, vol. 15, pp. 135–147, December 2014.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500